

# Encodage des textes patrimoniaux

## Transcription des Manuscrits

janvier 2012

# Cet objet, qu'est-ce que c'est ?



- ➊ Un texte !
- ➋ Un document !
- ➌ Un instant dans une série d'évènements !

## Un texte n'est pas un document

Un "document" est un objet existant dans le monde réel, qu'on peut numériser.

Un "texte" est une abstraction, la création d'une communauté de lecteurs, qu'on peut encoder.

## La théorie sousjacente

'In a perfect world, textual editing would be unnecessary' (Lance Schachterle)

Dans l'édition de sources les théories semblent inévitables :

- édition critique
  - "Lachmanienne"
  - "Bédieriste"
- édition diplomatique
- édition génétique

La TEI veut se présenter comme neutre et indépendante de toute théorie a priori... mais c'est une mission impossible...

# L'édition "Old Skool"

ACT II, SC. ii.]

OR, WHAT YOU WILL

103

She made good view of me, indeed so much,  
That me thought her eyes had lost her tongue, company with  
h 21  
For she did speake in starts distractedly.  
She loues me sure, the cunning of her passion  
Inuites me in this churlish messenger : 25  
None of my Lords Ring ? Why he sent her none ;  
I am the man, if it be so, as tis, 27

22. *That*] Knt, Coll. i, Hal. Dyce i,  
Cam. Sta. *That, as* Dyce ii, iii, Coll.  
iii, Huds. *That sure* Ff et cet.  
*me thought*] methought Rowe et  
seq. methought that Ktly.  
26. *none*;] *none*? Ff. *none*. Rowe.  
27. *I am the*] I should be Han.

27. *man, if it be so, as tis,*] *man, if*  
*it be so as tis,* Ff. *man—If it be so as*  
*'tis,* Rowe, Pope. *man—If it be so, (as*  
*'tis;)* Theob. Warb. Johns. Var. '73,  
'78. *man, if it be so; as 'tis,* Han. *man!*  
*If it be so,—as't is,—* Sta. *man:* *If it*  
*be so, (as 'tis)* Cap. et cet. (subs.)

22. *That*] The defective metre in this line is supposed to have been supplied by *sure* of the Second Folio. MALONE, however, was not altogether satisfied with it; he remarks that '*sure* is not very likely to have been the word omitted in the First

Une mise en page complexe, signalant la structuration du texte

- notes (plusieurs "étages")
- variantes
- références...

## Un exemple simple

Au dessous de ce vers de Hamlet :

LAERTES. Alas, then she is drowned.

se trouve un peu de discussion scientifique :

4.7.156 Alas, then is she drowned.] HIBBARD;  
Alas then, is she drown'd? F; Alas then is she drownd. Q3; Alas, then, she  
is drownd.  
Q2; So, she is drownde: Q1.

## Structuration classique d'apparat critique: <app>, <rdg>, et <lem>

- <app> (entrée d'apparat critique) contient une entrée dans un apparat critique, constituée d'un lemme facultatif et d'au moins une leçon.
- <rdg> (leçon) contient une leçon à l'intérieur d'une variante du texte.
- <lem> (lemme) contient le lemme ou le texte de base d'une variante textuelle.

## Par exemple ...

```
<app>
  <lem>Alas, then she is drowned.</lem>
  <rdg wit="#Hib">Alas, then is she drowned.</rdg>
  <rdg wit="#F">Alas then, is she drown'd?</rdg>
  <rdg wit="#Q3">Alas then is she drownd.</rdg>
  <rdg wit="#Q2">Alas, then, she is drownd.</rdg>
  <rdg wit="#Q1">So, she is drownde:</rdg>
</app>
```

# La granularité d'annotation n'est pas prédéfinie

```
<l n="156">
  <app>
    <rdg wit="#Hib">Alas, then</rdg>
    <rdg wit="#F">Alas then,</rdg>
    <rdg wit="#Q3">Alas then</rdg>
    <rdg wit="#Q2">Alas, then,</rdg>
    <rdg wit="#Q1">So,</rdg>
  </app>
  <app>
    <rdg wit="#Hib #F #Q3">is she</rdg>
    <rdg wit="#Q2 #Q1">she is</rdg>
  </app>
  <app>
    <rdg wit="#Hib">drowned.</rdg>
    <rdg wit="#F">drown'd?</rdg>
    <rdg wit="#Q3 #Q2">drownd.</rdg>
    <rdg wit="#Q1">drownde:</rdg>
  </app>
</l>
```

## Pointeurs et cibles

Les labels ('#Hib', '#Q1' etc.) sont des pointeurs: ils indiquent des métadonnées sur le témoin concerné

La TEI a beaucoup de propositions relatives à l'encodage des métadonnées... permettant ainsi l'évolution d'une 'codicologie quantitative'.

## La syntaxe peut être plus complexe...

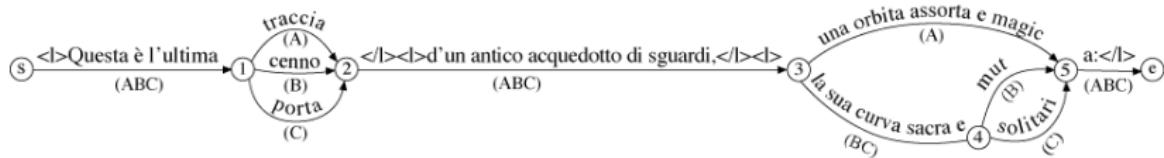
- Les lectures communes peuvent être regroupées avec `<rdgGroup>`, `<witList>` etc.
- La localisation des variantes représentées peut être représentée de plusieurs façons (par référence, par liens, etc.)
- L'existence d'un lemme n'est pas forcément obligatoire

Cependant, ce format ne suffit pas pour générer automatiquement les sources originales

Il ne serait pas non plus capable de représenter les sorties d'un logiciel de collation automatique des sources.

# Comment mieux modéliser la variation textuelle ?

- Un SIG du TEI MS a produit une revue critique de l'état actuel () .
- On peut aussi citer le modèle de Schmidt pour les 'multiversion documents':



- .. comme le "Rhine-delta model" de Sperberg-McQueen 1989 , cette modélisation serait mieux adaptée pour représenter les sorties d'un logiciel de collation automatique.

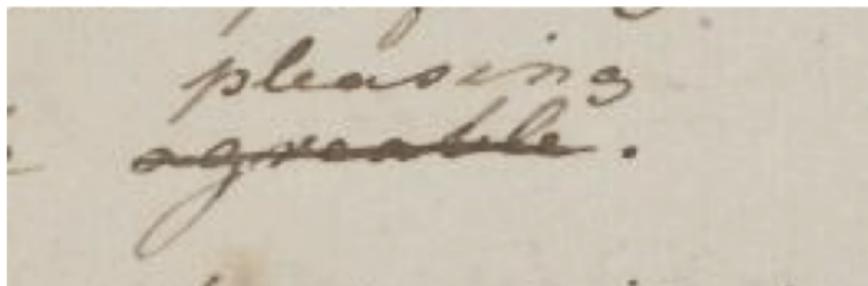
En effet, la représentation d'une collation déjà faite semble avoir moins d'intérêt que celle d'un dossier de transcriptions... un corpus

# Structures TEI pour la transcription des sources

- **<text>** : rassemble une lecture structurée du contenu intellectuel d'un document (ou en ensemble des documents) son 'texte'
- **<facsimile>** : regroupe un ensemble d'images représentant les pages (vel sim) d'un document
- **<sourceDoc>** : facsimile et transcription quasiment objective des aspects physique d'un seul document
- **<teiHeader>** : fournit des métadonnées qui décrivent les objets traités et les méthodes d'encodage utilisées

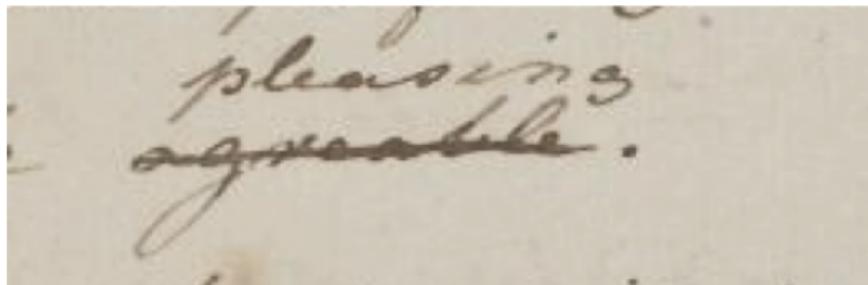
Une 'transcription' pourrait être considérée soit comme un 'text',  
soit comme un 'document'

## Comment encoder ceci ?



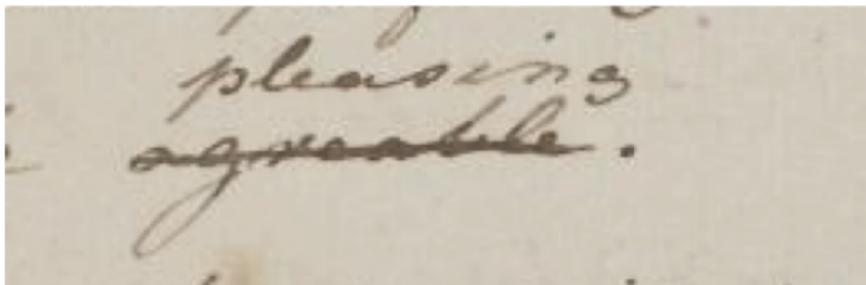
- ① 'agreeable' est raturé, 'pleasing' est écrit au dessus, dans un même espace.
- ② 'agreeable' est supprimé et remplacé par 'pleasing'
- ③ Le texte original était 'agreeable', mais quelques temps après, ce mot a été supprimé. Le mot 'pleasing' a été ajouté dans ce même contexte.

## Comment encoder ceci ?



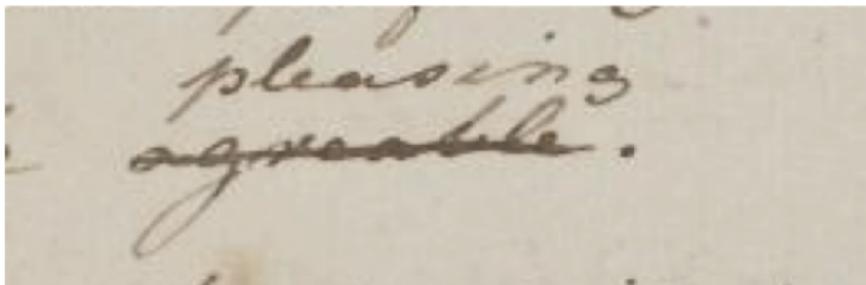
- ➊ 'agreeable' est raturé, 'pleasing' est écrit au dessus, dans un même espace.
- ➋ 'agreeable' est supprimé et remplacé par 'pleasing'
- ➌ Le texte original était 'agreeable', mais quelques temps après, ce mot a été supprimé. Le mot 'pleasing' a été ajouté dans ce même contexte.

## Comment encoder ceci ?



- ➊ 'agreeable' est raturé, 'pleasing' est écrit au dessus, dans un même espace.
- ➋ 'agreeable' est supprimé et remplacé par 'pleasing'
- ➌ Le texte original était 'agreeable', mais quelques temps après, ce mot a été supprimé. Le mot 'pleasing' a été ajouté dans ce même contexte.

## Comment encoder ceci ?



- ① 'agreeable' est raturé, 'pleasing' est écrit au dessus, dans un même espace.
- ② 'agreeable' est supprimé et remplacé par 'pleasing'
- ③ Le texte original était 'agreeable', mais quelques temps après, ce mot a été supprimé. Le mot 'pleasing' a été ajouté dans ce même contexte.

# Transcription: une forme de lecture

Quel est le but de votre transcription?

- rendre accessible la source ...
- ... et donc compréhensible
- ce qui implique peut être l'ajout (ou la suppression) des infos supplémentaires

Donc,

- toute transcription implique une sélection
- toute transcription implique de l'imagination
- toute transcription implique des interprétations

## La transcription des sources primaires

Quelles caractéristiques d'une source primaire pourrait-on vouloir exprimer dans un travail de transcription ?

- la manière dont le texte est inscrit sur la page
- l'orthographe, le rendu formel original, la ponctuation originale
- les abréviations
- les ajouts et suppressions
- les erreurs et omissions 'évidentes'
- éventuellement, les illustrations, dessins ou graphiques illustrant le texte
- les passages endommagés, ou illisibles, à restaurer.

## Encodage de la structure physique

Pour exprimer la manière dont le texte est inscrit sur le support physique, des éléments vides de type 'milestone' (borne) sont disponibles, `<gb/>`, `<pb/>`, `<cb/>`, et `<lb/>`, qui correspondent respectivement à la désignation des débuts de cahier ("gathering"), page, colonne et ligne.

Ceux-ci servent principalement à établir un système de référencement.

L'élément `<fw>` sert à baliser le paratexte; l'élément `<handShift>` les changements d'écriture.

# Un encodage minimal typique

5  
demy faburcans &c.  
soixante sept pieces au marc  
a ung felin et demy de remede  
cours chacune piece pour po  
cinquante soubz tournois.  
  
Sur le pris de quinze livres  
tournois le marc dargent le roy  
fust continuee la fabricacion  
des testons a dix deniers  
de six grains troyens quartz  
et fin qui valent unze deniers  
six grains dargent le roy a

<p>  
<!-- ... -->  
<pb n="5r"/><fw>5</fw>  
<lb><expan>et</expan> demy trebuchans de  
<lb>soixante sept pieces au marc  
<lb>a ung felin <expan>et</expan> demy de  
remede  
<lb>Cours chacune piece  
<expan>pour</expan>  
<lb><expan>cinquante</expan> soubz  
<expan>tournois</expan> <pc>.</pc>  
</p> <p>  
<lb>Sur le pris de quinze livres  
<lb><expan>tournois</expan> le  
marc dargent le roy  
<lb>fust <expan>continuee</expan> la  
<expan>fabricacion</expan>  
<lb>des testons a dix deniers  
<lb>dix huict grains troyens quartz  
<lb>de fin qui <expan>valent</expan> unze  
deniers  
<lb>six grains dargent le roy<pc>.</pc> a

## Corrections

L'élément `<sic>` peut être utilisé pour indiquer que la leçon donnée par le manuscrit est erronée ou n'a pas de sens, tandis que l'élément `<corr>` sert à fournir ce qui est la leçon correcte selon l'opinion de l'éditeur :

```
<sic>relea</sic>
```

```
<corr>relicta</corr>
```

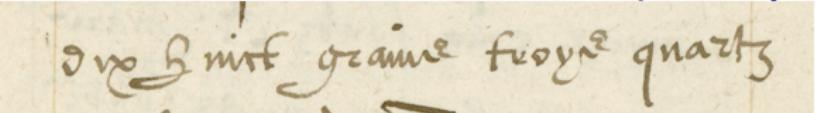
Les deux éléments peuvent être associés au sein d'un élément `<choice>` :

```
<choice>
  <sic>relea</sic>
  <corr>relicta</corr>
</choice>
```

## Normalisation

Les sources primaires utilisent rarement l'orthographe moderne. Pour la recherche ou pour d'autres raisons liées aux traitements informatiques prévus, la graphie moderne peut s'avérer utile dans une transcription. L'élément `<reg>` (régularisation) est disponible pour encoder une forme normalisée, tandis que l'élément `<orig>` (forme originale) contient la graphie d'origine non normalisée. On peut, si on le souhaite, regrouper ces éléments de contenu alternatif en utilisant l'élément `<choice>`.

## Normalisation (exemple)



dix huit graine troye quartz

```
<lb/>dix <choice>
<orig>huict</orig>
<reg>huit</reg>
</choice> grains
<choice>
<orig>troys
quartz</orig>
<reg>trois-quart</reg>
</choice>
```

Dans ce cas, il existe aussi la possibilité d'une régularisation sémantique:

```
<lb/>
<measure quantity="18.75" unit="gr">dix
<choice>
<orig>huict</orig>
<reg>huit</reg>
</choice> grains <choice>
<orig>troys
quartz</orig>
<reg>trois-quart</reg>
</choice>
</measure>
```

## Les abréviations

Les abréviations sont très caractéristiques des manuscrits de toutes sortes. On distingue plusieurs types d'abréviations :

**Abréviations par suspension** la ou les premières lettres du mot sont écrites, suivies en général d'un point ou d'une autre marque : par exemple 'e.g.' pour 'exempli gratia'.

**Abréviations par contraction** les première(s) et dernière(s) lettres du mot sont écrites, accompagnées le plus souvent d'un signe abréviaatif tel que trait suscrit, ou, plus rarement, un point : par exemple 'ā' pour 'an', 'M.' pour 'Monsieur' etc.

**"Brévigraphes"** ce sont des symboles spécifiques, tels que la note tironienne (&#x204A; ☐) utilisée à la place de la conjonction de coordination 'et', la lettre p barrée (&#x1D71; ☐), souvent utilisée à la place de la syllabe 'per'.

La plupart des symboles dont on a besoin sont définis par Unicode, mais ne sont pas forcément présents dans toutes les polices.

## Les abréviations et leur résolution

Une abréviation peut être restituée de deux manières :

- on peut choisir de consigner l'abréviation non développée, en la transcrivant simplement telle quelle : par exemple, 'po suivi de r suscrit' ou un 'a avec macron'
- on peut aussi interpréter ou développer l'abréviation, en remplaçant la ou les lettres par leur signification : par exemple, 'pour', 'an'

La TEI permet de fournir à la fois les formes abrégées et développées des abréviations.

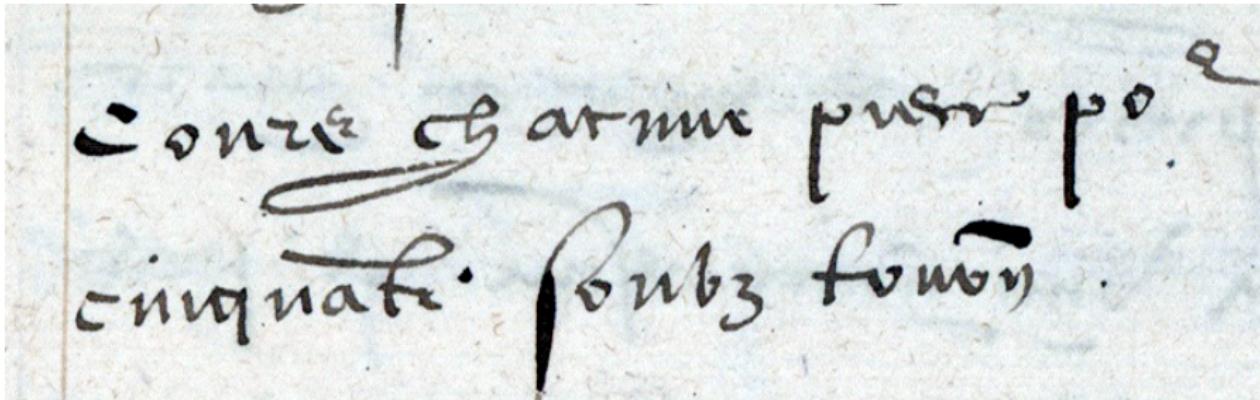
## Encoder les abréviations

La TEI propose deux niveaux d'encodage :

- la totalité d'un mot abrégé et la totalité de sa forme développée peuvent être encodés au moyen des éléments `<abbr>` et `<expan>`
- le symbole utilisé pour indiquer la suppression d'une ou plusieurs lettres peut être encodé par l'élément `<am>`; les lettres substituées à ce symbole lorsqu'on résout l'abréviation par l'élément `<ex>`

Ici aussi on peut utiliser les deux niveaux conjointement.

## Un exemple simple (1)



Cours chacune piece pour  
cinquante soubz tournois.

Selon la stratégie éditoriale, on peut signaler simplement qu'il y avait des formes abrégées qu'on a rétablies:

```
<p>
<lb/>Cours chacune piece <expan>pour</expan>
<lb/>
<expan>cinquante</expan> soubz <expan>tournois</expan>
<pc>.</pc>
</p>
```

## Exemple simple (2)

Mais, en effet, ‘pour’ a été écrit ‘po’ suivi d'une lettre ‘r’ suscrite ; ‘cinquante’ comme ‘cinquāte’ avec un macron sur le ‘a’ pour signaler la nasalisation.

On pourrait donc encoder ces mots de l'une ou l'autre des manières ci-après :

```
<p>
  <abbr>po&#xFFD; </abbr> .... <abbr>cinquāte</abbr>
</p>
```

... OU :

```
<p> po<am>&#xFFD; </am> ... ou po<ex>u</ex>r </p>
```

```
<p>
  <abbr>po<am>&#xFFD; </am>
  </abbr> ou <expan>po<ex>u</ex>r</expan>
</p>
```

## Exemple simple (3)

Et, bien sûr, la TEI permet d'avoir et le beurre et l'argent du beurre

```
<p> po<choice>
  <am>&#FFFD;</am>
  <ex>ur</ex>
</choice>
</p>
```

```
<choice>
  <abbr>po<am>&#xFFD;</am>
  </abbr>
  <ex>ur</ex>
</choice>
```

## Catégoriser les abréviations

L'attribut @type de l'élément `<abbr>` sert à catégoriser les abréviations, qu'il s'agisse de préparer une analyse statistique ou de restituer ces abréviations de manière différente selon leur type.

Par exemple, on peut vouloir afficher le développement des abréviations par suspension entre crochets, d'autres types en italique ou entre parenthèses etc.

```
<choice>
  <abbr type="brevigraphe">po<am>xFFFF;</am>
  </abbr>
  <expan>po<ex>u</ex>r</expan>
</choice>
```

L'inscription encodée ci-dessus peut être affichée comme suit :  
'po(u)r'

Comme ailleurs, les attributs `@resp` et `@cert` sont disponibles pour indiquer l'agence responsable de la forme développée, et le degré de certitude de cette interprétation.

## Ajouts, suppressions et substitutions

Les changements opérés sur le texte inscrit, qu'ils soient faits par le scripteur ou ultérieurement par une autre main, peuvent être encodés au moyen des éléments `<add>` (ajout) ou `<del>` (suppression).

- `<add>` contient des lettres, des mots ou des phrases insérées dans le texte par un auteur, un copiste, un annotateur ou un correcteur.
- `<del>` contient une lettre, un mot ou un passage supprimé, marqué comme supprimé, sinon indiqué comme superflu ou erroné dans le texte par l'auteur, par un copiste, un annotateur ou un correcteur.

Lorsque les ajouts et suppressions sont considérés comme une seule opération de substitution, les éléments `<add>` et `<del>` peuvent être groupés au moyen de l'élément `<subst>` (substitution).

## Substitutions

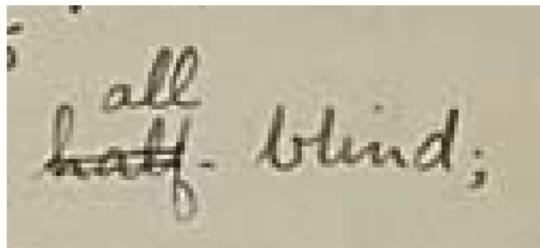
L'élément `<subst>` (substitution) associe une ou plusieurs suppressions avec un ou plusieurs ajouts lorsque cette combinaison peut être considérée comme une seule intervention sur le texte.

Exemples :

- un mot ou un caractère écrit par-dessus un autre
- un mot ou un caractère supprimé, remplacé par un autre écrit au-dessus par la même main lors de la rédaction du texte
- un mot ou un caractère supprimé, remplacé par une main différente à un autre moment

`<subst>` ne suffit pas pour l'encodage des cas plus complexes, par exemple une longue série de substitutions dont l'ordre peut être incertain, ainsi que le choix de la leçon à préférer.

## (Sans doute) une substitution

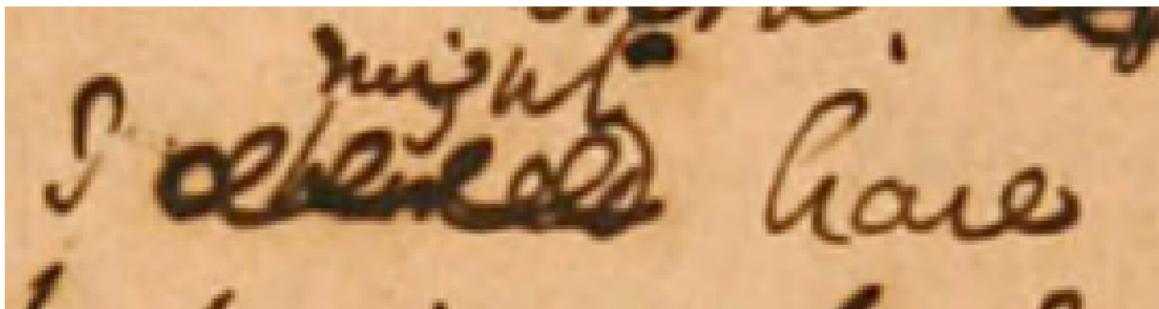


```
<subst>
  <del>half-</del>
  <add>all</add>
</subst> blind
```

## Difficultés de lecture

Utiliser l'élément `<unclear>` si le texte peut être lu mais sans certitude.

Utiliser l'attribut `@reason` pour déclarer la cause de l'incertitude dans la transcription et l'attribut `@cert` pour donner une idée de la confiance que l'on peut avoir dans la transcription.



```
I <subst>
<add>might</add>
<del>
  <unclear reason="overinking" cert="medium" resp="#LDB">should</unclear>
</del>
</subst>have
```

## Dommages matériels

L'élément `<damage>` est à utiliser pour indiquer les zones du document sur lesquelles des dommages matériels affectent le texte, mais où l'on peut lire au moins une partie du texte avec certitude. Les attributs `@agent` et `@extent` précisent respectivement la cause et l'étendue des dommages.

```
<l>The Moving Finger wri<damage agent="water" group="1">es; and</damage>
having writ,</l>
<l>Moves <damage agent="water" group="1">
  <supplied>on: nor all your</supplied>
</damage>
Piety nor Wit</l>
```

## Lacunes

Quand il est impossible de fournir une lecture, on se sert de l'élément `<gap>` : son attribut `@reason` fournit une explication de la lacune, et ses attributs `@extent` et `@unit` indique sa taille.

```
<gap reason="wormhole" extent="7" unit="mm"/>
```

I am dr Sr yr <gap reason="illegible" quantity="3" unit="word"/>Sydney  
Smith

## Un exemple plus complet

And towards our distant rest began to trudge,  
~~Helping the worst amongst us~~, who'd no boots all  
But limped on, blood-shod. All went lame; ~~half-~~ blind;  
Drunk with fatigue ; deaf even to the hoots  
Of tired, outstripped ~~fif~~ five-nines that dropped behind.

```
<l>And towards our distant rest began to trudge,</l>
<l>
<subst>
  <del>Helping the worst amongst us</del>
  <add>Dragging the worst amongst us</add>
</subst>, who'd no boots
</l>
<l>But limped on, blood-shod. All went lame; <subst>
  <del status="shortEnd">half-</del>
  <add>all</add>
</subst> blind;</l>
<l>Drunk with fatigue ; deaf even to the hoots</l>
<l>Of tired, outstripped <del>fif</del> five-nines that dropped behind.</l>
```

## Quelques problèmes

Ces méthodes sont adéquates à des situations de variation pas trop complexes. On peut néanmoins rencontrer quelques problèmes quand...

- il y a des chevauchements (et il y en a toujours)
- on est concerné par la séquence des interventions
- la signification et la mise en forme sont difficiles à distinguer

Il y a des astuces bien sur ...

## Interventions qui traversent les frontières des éléments

Les éléments `<addSpan>` et `<delSpan>` délimitent une section de texte au moyen de pointeurs plutôt qu'en encadrant une section de texte. Ils sont utiles si un ajout ou une suppression chevauche une autre portion de texte.

L'attribut `@spanTo` indique où se situe la fin d'une portion de texte commençant par l'élément qui porte l'attribut.

```
<l> .... <delSpan spanTo="#id4"/> ... </l>
<l> ... <anchor xml:id="id4"/> ... </l>
<l> .... </l>
```

## Usage des attributs pour capturer la séquence

- L'auteur (WJ) a écrit 'One must have lived... '
- Il rajoute un mot 'But' avant 'One'
- L'éditeur (FB) corrige 'One' en 'one'

```
<add place="supra" hand="#WJ" cert="medium"> But</add>
<choice>
  <sic>One</sic>
  <corr resp="#FB" cert="high">one</corr>
</choice> must have lived ...
<!-- elsewhere --&gt;
&lt;respStmt xml:id="FB"&gt;
  &lt;resp&gt;editorial changes&lt;/resp&gt;
  &lt;name&gt;Fredson Bowers&lt;/name&gt;
&lt;/respStmt&gt;
&lt;respStmt xml:id="WJ"&gt;
  &lt;resp&gt;authorial changes&lt;/resp&gt;
  &lt;name&gt;William James&lt;/name&gt;
&lt;/respStmt&gt;</pre>
```

## La balise <restore> peut indiquer un changement d'avis

- On écrit 'For I hate this my body'
- Le mot 'my' est supprimé
- On écrit dans le marge 'stet'

```
<l>[...] For I hate this <restore hand="#dhl" type="marginalStetNote">
<del>my</del>
</restore> body [...]</l>
```

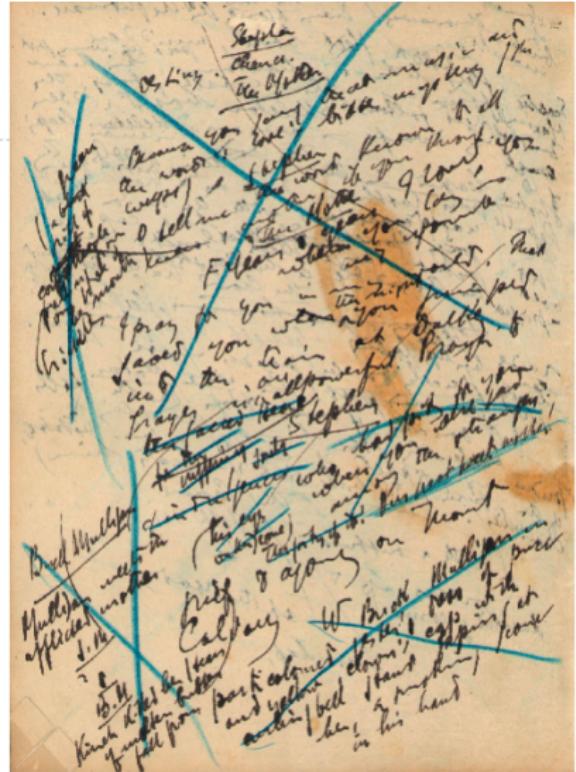
... notez que nous n'avons pas saisi le 'metamark' 'stet', bien que nous ayons indiqué sa portée.

## Et les manuscrits modernes ?

"manuscrits qui font partie d'une genèse textuelle attestée par plusieurs témoins successifs et qui manifestent le travail d'écriture d'un auteur" (Almuth Grésillon)

- "an author is not a scribe" (Edward Vanhoutte)
- les fonctions des objets manuscrits modernes et anciens ne sont pas forcément pareilles...

## Par exemple...



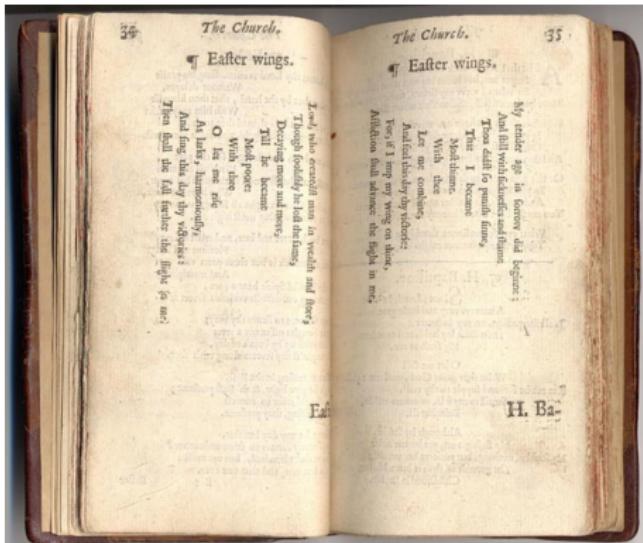
## La TEI fournit un moyen d'encoder des **textes**, mais ..

- Le processus de création du document est au moins aussi important que le texte qu'il contient
- Le texte n'a pas de sens indépendamment du document : que ce soit parce que
  - le texte est présenté graphiquement
  - le texte est non-linéaire, aléatoire, combinatoire....

Les textes non-linéaires sont des non-textes (Daniel Ferrer)

## Text/Image

A toute époque on trouve des documents 'ludiques', dont la signification dérive en partie de sa mise en forme, plutôt que de ses propriétés linguistiques, ou de la tension entre les deux



D'abord, la TEI considérait ce genre de problèmes hors de sa portée, en partie à cause des problèmes techniques.

## Encodage documentaire

Nous avons déjà évoqué les problèmes associés à l'encodage de

- l'apparence physique d'un document
- la disposition des unités structurant un texte sur les pages physiques d'un document
- la séquence de composition des morceaux d'un document
- les indications 'meta' d'un écrivain sur la lecture voulue de ses écrits

La TEI d'aujourd'hui propose des mécanismes plus ou moins élaborés pour l'encodage de ces phénomènes...

## Qu'est-ce qu'un "facsimilé numérique"?

- Au plus simple, un inventaire de fichiers images, qui reproduisent le document décrit par l'en-tête associé.
- mais peut également définir des espaces visuels plus petits, comme les surfaces ou zones d'intérêt sur ces surfaces
- Tous peuvent être alignés, à un niveau approprié, avec les parties d'une transcription

# La solution la plus simple pour réaliser un "facsimilé numérique"

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- -->
  </teiHeader>
  <text>
    <pb facs="page1.png"/>
    <!-- texte de la page 1 facultativement transcrit ici -->
    <pb facs="page2.png"/>
    <!-- texte de la page 2 facultativement transcrit ici-->
  </text>
</TEI>
```

## Inconvénients...

- difficile de supporter des relations plus compliquées
- maintien difficile d'informations propres à l'image
- nécessite donc l'intégration avec d'autres fichiers  
(typiquement METS)

## Solution un peu moins simple

Regrouper les données à un seul endroit:

```
<facsimile  
    xml:base="http://mylibrary.wibble.fr/blah/blah/blah">  
    <graphic url="page1.png"/>  
    <graphic url="page2.png"/>  
    <graphic url="page3.png"/>  
    <graphic url="page4.png"/>  
</facsimile>
```

(On pourrait également se servir de [\*\*<binaryObject>\*\*](#))

## Et si on disposait de versions alternatives d'une même image ?

Se servir de <surface>pour grouper les images équivalentes :

```
<facsimile>
  <graphic url="page1.png"/>
  <surface>
    <graphic url="page2-highRes.png"/>
    <graphic url="page2-lowRes.png"/>
  </surface>
  <graphic url="page3.png"/>
  <graphic url="page4.png"/>
</facsimile>
```

# Lier l'image à la transcription

```
<facsimile>
  <graphic xml:id="page1" url="page1.png"/>
  <graphic xml:id="page2" url="page2.png"/>
  <graphic xml:id="page3" url="page3.png"/>
  <graphic xml:id="page4" url="page4.png"/>
</facsimile>
<!-- ... -->
<text>
  <pb facs="#page1"/>
<!-- texte de la page 1 facultativement transcrit ici -->
  <pb facs="#page2"/>
<!-- texte de la page 1 facultativement transcrit ici -->
</text>
```

## Lier deux parties d'une image

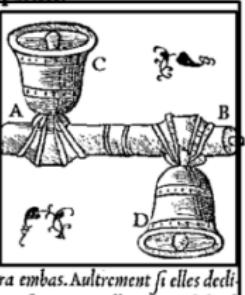
- on se sert de `<surface>` pour définir un système abstrait de coordonnées
- on se sert de `<zone>` pour définir une zone d'intérêt
- on se sert de `<graphic>` pour fournir une image de la zone ou surface qui le contient

**EDV SON ET ACCORD DES CLOCHES ET**  
des alleurees des cheualx, chariotz & charges, des fontaines: &  
encyclie du monde, & de la dimension du corps humain.

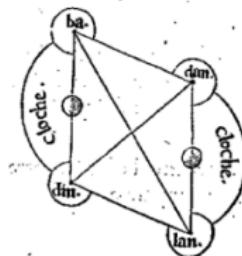
Chapitre septiesme.

1. **Le son & accord des cloches pendans en vng mesme axe, est fait en contraires parties.**

**L**es cloches ont quasi figures de rondes pyramides imperfæctes & irregulieres: & leur accord se fait par regle geometrique. Comme si les deux cloches C & D sont pendans à vng mesme axe on eſſeu A B: ie dis que leur accord se fera en contraires parties, cōne voyes icy figuré. Car quid lune ſera en hault, lautre declinera embas. Autrement ſi elles dediuent toutes deux enſemble en vne meſme partie, elles feront diſcord & ſera leur ſonnerie mal plaiſante à oyr.



2. **Le vray accord de deux cloches par lattouchement des bataulx, est fait diametralement.**



**C**haſcune cloche a deux diuers coſtez, entre leſ quels le batail touche par diuerſes fois. Et à caufe que l'accord de deux cloches ſe fait par l'inclination delles en diuerses & contraires parties: Il faut que ledit accord ſe fasse diametralement G.i.

# Encodage pour cette image

```
<facsimile>
  <surface ulx="0" uly="0" lrx="200" lry="300">
    <zone xml:id="B49r" ulx="0" uly="0" lrx="200" lry="300">
      <graphic url="Bovelles-49r.png"/>
    </zone>
    <zone ulx="105" uly="76" lrx="175" lry="160">
      <graphic url="Bovelles49r-detail.png"/>
    </zone>
    <zone xml:id="B49rHead" ulx="25" uly="25" lrx="180" lry="60"/>
  <!-- contient le titre -->
  <zone xml:id="B49rPara2" ulx="28" uly="75" lrx="175" lry="178"/>
  <!-- contient le paragraphe en italique -->
  <zone xml:id="B49rFig1" ulx="105" uly="76" lrx="175" lry="160"/>
  <!-- contient l'image de cloche -->
  <zone xml:id="B49rW457" ulx="45" uly="125" lrx="60" lry="130"/>
  <!-- contient le mot "pendans" -->
  </surface>
</facsimile>
```

(Notons que nous introduisons des identifiants pour chaque zone)

# La transcription

```
<pb  facs="#B49r"/>
<fw>De Geometrie 49</fw>
<head  facs="#B49rHead"> DU SON ET ACCORD DES CLOCHES ET <lb/> des
alleures des chevaux, chariotz & charges, des fontaines:& <lb/> encyclie du
monde, & de la dimension du corps humain.</head>
<head>Chapitre septiesme</head>
<div  n="1">
  <p>Le son & accord des cloches pendans en ung mesme <lb/> axe, est faict en
  contraires parties.</p>
  <p  rend="it"  facs="#B49rPara2">LEs cloches ont quasi fi<lb/>gures de rondes
  pyra<lb/>mides imperfaictes & <lb/> irregulieres: & leur accord se <lb/> fait
  par reigle geometrique. Com<lb/>me si les deux cloches C & D <lb/> sont
  <w  facs="#B49rW457">pendans</w> à ung mesme axe <lb/> ou essieu A B: je dis que leur
  ac<lb/>cord se fera en co<ex>n</ex>traires parties<lb/> co<ex>m</ex>me voyez icy
  figuré. Car qua<ex>n</ex>d <lb/> lune sera en hault, laultre declinera embas.
  Autrement si elles decli<lb/>nent toutes deux ensembles en une mesme partie, elles
  seront discord, <lb/> & sera leur sonnerie mal plaisante à oyr.<figure  facs="#B49rFig1"/>
</p>
</div>
```

# Comment combiner transcription et images ?

On peut se servir de l'élément `<sourceDoc>`...

- au même niveau que `<text>` et `<teiHeader>`
- représente la **structure physique** du document
- rassemble un ou plusieurs éléments `<surface>` écrites, éventuellement divisées en `<zone>`s d'intérêt
- ou en `<line>`s topographiques

Samaria is a Greek  
Gorge of water that  
comes from the natural  
springs of Stilos, in  
Crete.

1 April 2009

Fed Birds in the park today.  
Might write an article about  
the Thick-billed Warbler.

Samaria is a Greek  
branch of water that  
comes from the natural  
springs of Stilos, in  
Crete.

1 April 2009

Feed Birds in the park today.  
Might write an article about  
the Thick-billed Warbler.

## Encodage "textuel"

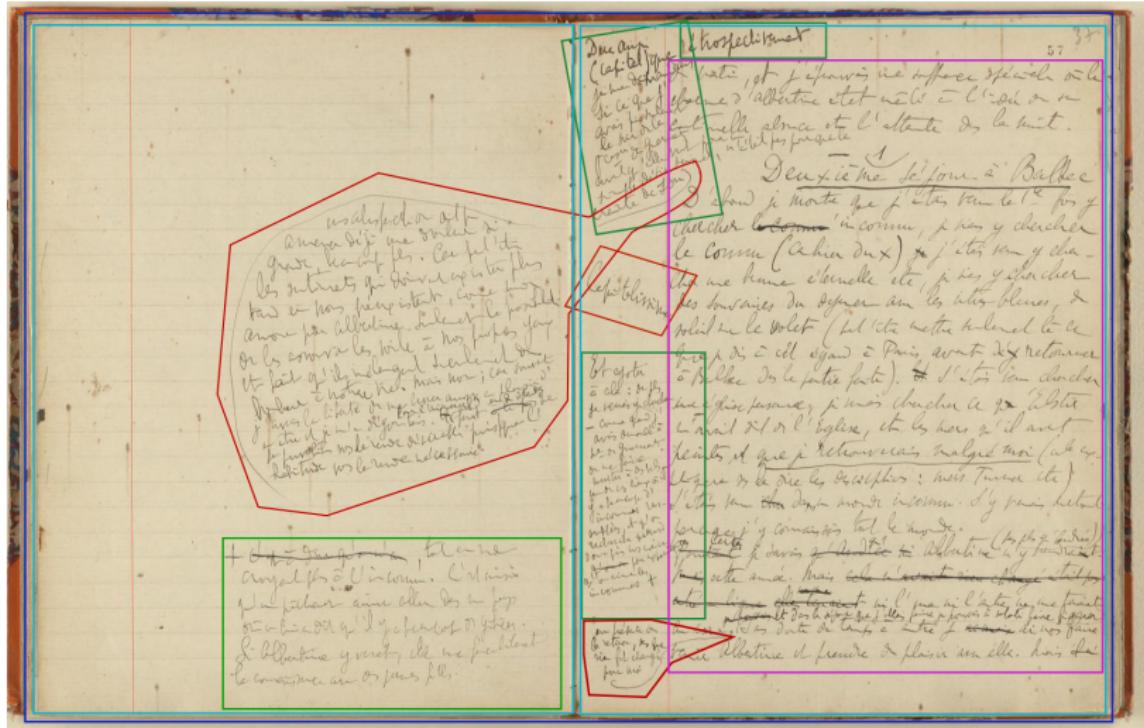
```
<text>
  <body>
    <div type="diary-entry">
      <dateline>
        <date value="2009-04-01">1 april 2009</date>
      </dateline>
      <p>
        <lb/>Fed Birds in the park today. <lb/> Might write an article about
      <lb/>the
        Thick-billed Warbler.</p>
    </div>
    <div type="note" rend="rotated">
      <p>
        <lb/>Samaria is a Greek<lb/>brand of water that<lb/>comes from the
        natural<lb/>springs of Stilos, in <lb/>Crete.</p>
    </div>
  </body>
</text>
```

## Encodage "documentaire"

```
<sourceDoc>
  <surface uly="0" lry="300">
    <zone uly="43" lry="84">
      <line rend="right">1 April 2009</line>
    </zone>
    <zone>
      <line>Fed Birds in the park today.</line>
      <line>Might write an article
          about</line>
      <line>the Thick-billed Warbler.</line>
    </zone>
    <zone uly="20" lry="60" rotate="90">
      <line>Samaria is a
          Greek</line>
      <line>brand of water that </line>
      <line>comes from the natural
          </line>
      <line>springs of Stilos, in </line>
      <line>Crete.</line>
    </zone>
  </surface>
</sourceDoc>
```

## Les surfaces et les zones...

La relation surface / zone peut être assez complexe :



## Surfaces multi-parties

... très complexe

[2] |

Urt 15 Avril 79

1

la Communauté  
du Mag

ideologique

Depuis longtemps, je m'intéressais à la Photographie. J'admirais contre le cinéma, dont j'étais arrivé à saisir, à tout prix, le trait fondamental de représentation. Bref, à la Photographie, sa façon de fixer rien même, mais quelque chose qui l'est lié à quelque chose d'autre, quelque chose d'étrange, d'extraordinaire, par une sorte de charme, d'expression pénétrante, parfois sombreuse jusqu'au déchirement.

\* Une telle œuvre suffit dire qu'en fond, en dehors de toutes sortes de l'espace, je l'avais pas à être sûr que la Photographie existait.

2

qui pouvait me guider

Dès le premier pas scientifique, celui du classement (il faut bien classer, échantillonner, si l'on veut constituer un corpus), la Photographie se dérobait, les répartitions auxquelles on la faisait empiriques (Professionnels, Amateurs), l'institution,

## Balisage des informations "meta" ...

<metamark> une marque quelconque déterminant la lecture du document

<mod> et <modSpan> indiquent, de manière générique, le lieu d'une modification dans le document, sans spécifier sa fonction

<undo> <redo>, <retrace> indications respectivement de : corrections supprimées, corrections réaffirmées, réécritures

<transpose> et <transposeGrp> transpositions

## Metamark: exemple

I am that halfgrown <sup>aw<sup>s</sup></sup> boy fallen asleep,  
The tears of foolish passion yet undried  
Upon my cheeks.

Years with all their events pass for me,  
Some are spent in travel.—Some in the ~~usual~~  
usual hunt ~~for~~ <sup>after</sup> fortune.

I pass through <sup>the</sup> travels and fortunes of those  
years and become old,  
Each in its due order comes and goes,  
And then a message for me comes.

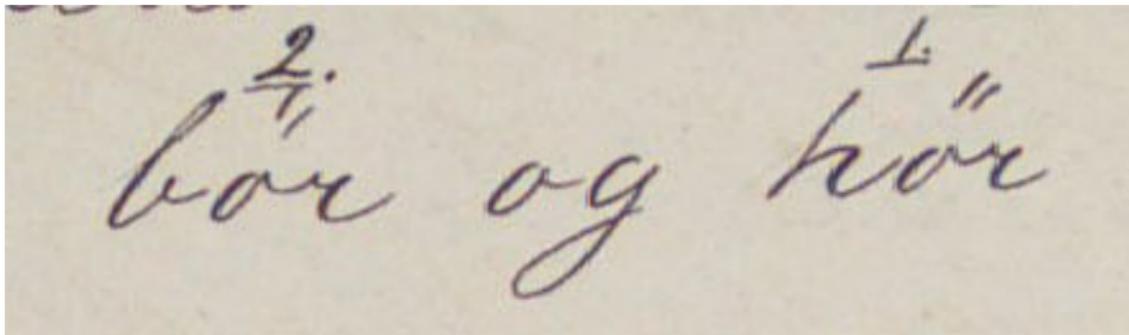
The

~~Entered~~ - YG  
==

## Metamark : exemple encodé

```
<surface>
  <metamark function="used" rend="line" target="#X2"/>
  <zone xml:id="X2">
    <line>I am that halfgrown <add>angry</add> boy, fallen asleep</line>
    <line>The tears of foolish passion yet undried</line>
    <line>upon my cheeks.</line>
  <!-- ... -->
    <line>I pass through <add>the</add> travels and <del>fortunes</del> of
    <retrace>thirty</retrace>
    </line>
    <line>years and become old,</line>
    <line>Each in its due order comes and goes,</line>
    <line>And thus a message for me comes.</line>
    <line>The</line>
  </zone>
  <metamark function="used" target="#X2">Entered - Yes</metamark>
</surface>
```

## <Transposition> : exemple



Extrait de

```
<line>
  <seg xml:id="ib01">bör</seg>
  <metamark rend="underline" function="transposition" target="#ib1" place="above">2.</metamark> og <seg xml:id="ib02">hör</seg>
  <metamark rend="underline" function="transposition" target="#ib02" place="above">1.
  </metamark>
</line>
```

## Transposition : exemple

```
<transposeGrp>
  <transpose>
    <ptr target="#ib02"/>
    <ptr target="#ib01"/>
  </transpose>
</transposeGrp>
```

## Distinction des étapes d'écriture

L'attribut @change sert à associer n'importe quel élément de la transcription avec une étape identifiée dans l'évolution du document

L'identification d'une étape est au cœur de la problématique génétique : on ne peut pas vous aider là !

Les étapes elles-mêmes sont à documenter dans l'élément `<profileDesc>` de l'en-tête TEI.

## Example...

Ex hypothese, les couleurs d'encre sont à associer à des niveaux distincts d'écriture.

# Documentation des étapes

```
<profileDesc>
  <creation>
    <listChange ordered="true">
      <change xml:id="ST-1">Premier brouillon, à l'encre noire </change>
      <change xml:id="ST-2">Corrections et révisions en rouge</change>
      <change xml:id="ST-3">Corrections et révisions en bleu</change>
      <change xml:id="ST-4">Ratures et indications d'usage en vert</change>
    </listChange>
  </creation>
</profileDesc>
```

# Association transcription/documentation

```
<zone xml:id="zone1" change="#ST-1">
    <line> 28) le court de tennis. Les tribunes sont ... Deux joueurs</line>
<!-- ... -->
    <line>30) l'un des joueurs de tennis se tient ... trois</line>
    <line>fois sur le sol</line>
<zone change="#ST-2">
    <line>31) </line>
    <line>Vue de face</line>
    <line> à contre jour</line>
    <metamark function="add"/>
    <line>la vieille dame ... dans le vestibule (contre-jour)</line>
</zone>
<line>3<subst change="#ST-2">
    <del>1</del>
    <add>2</add>
    </subst>)Le groupe de cavaliers
    ...</line>
<!-- ... -->
    <line>... Le mot FIN apparaît sur l'écran.</line>
</zone>
<zone change="#ST-3">
    <line>Dans ce cas...</line>
</zone>
```